

## Taller 2: Machine Learning con Python

Profesor: Dr. Ing. Rodrigo Salas

Machine Learning con Python

## 2.1 Descripción de la Tarea

En este trabajo se espera que el estudiante realice un estudio comparativo del desempeño de las técnicas. Para ello Ud. deberá aplicar Cross-Validation con 10-fold y obtener la media y desviación estándar de las métricas de desempeño. Construir una tabla comparativa de los métodos.

Deberá entregar el archivo Jupyter y un archivo pdf con las tablas resultantes con una discusión de los resultados

Para realizar el estudio experimental deberá utilizar los siguientes conjuntos de datos extraídos de la UCI Machine Learning Repository

<https://archive.ics.uci.edu/ml/index.php>

- Hepatitis <https://archive.ics.uci.edu/ml/datasets/Hepatitis>
- Pima Indians Diabetes <http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>
- Water Treatment Plant. <http://archive.ics.uci.edu/ml/datasets/Water+Treatment+Plant>

## 2.2 Actividades

1. Describir los conjuntos de datos, con sus atributos y el problema que se espera resolver.
2. Escoger 3 tipos de técnicas de clustering para ser implementadas con los datos. Entregar una tabla comparativa del índice de Silhouette las 3 técnicas de cluster utilizadas.
3. Escoger 3 tipos de técnicas de Clasificación para ser implementadas con los datos. Entregar una tabla comparativa del *accuracy*, *precision* y medida *F* de las tres técnicas de clasificación escogidas.
4. Repetir el paso anterior, pero aplicando la técnica de PCA utilizando las primeras componentes principales que justifiquen a lo menos el 70% de la varianza.

Consideraciones:

- Fecha de entrega: 8 de Diciembre del 2017 a las 23:55 utilizando la plataforma moodle
- Cada día de atraso será penalizado con 20 puntos.
- Pueden discutir los problemas con sus compañeros, sin embargo la entrega y el trabajo es de grupos de 2 a 3 estudiantes máximo.

## References

- [1] Scikit-learn.org, *Machine Learning in Python*. <http://scikit-learn.org/>
- [2] ScyPy.org, *Statistical Functions (scipy.stats)*. <http://docs.scipy.org/doc/scipy/reference/stats.html>
- [3] Pandas, *Python Data Analysis Library*. <http://pandas.pydata.org/>
- [4] Numpy, *Fundamental package for scientific computing with Python*. <http://www.numpy.org/>
- [5] Matplotlib, *Biblioteca gráfica para python*. <http://matplotlib.org>